

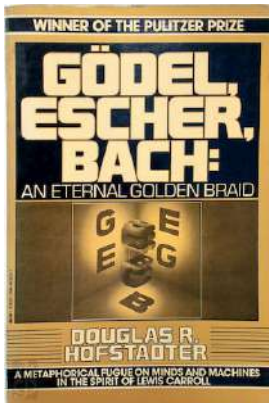
Wiskunde, kunstmatige intelligentie en vreemde lussen

door [Bart Verheij](#) (Head of department AI, chair of AI & argumentation)

Sinds de lancering van ChatGPT (november 2022) gaat het nog meer dan daarvoor over de kansen en gevaren van AI en hoe AI binnenkort alle menselijke taken gaat overnemen. Is dat eigenlijk zo? Blijft er een plek voor de mens?

Vanwege een lezing voor wiskundigen (op het Wintersymposium van het Koninklijk Wiskundig Genootschap op 14 januari 2023 [1]), vroeg ik me af hoe goed het systeem eigenlijk was in wiskunde. Ik dacht terug aan mijn tijd als wiskundestudent in Amsterdam (UvA) toen ik bij professor Gerard van der Geer afstudeerde in de algebraïsche meetkunde [2], en vroeg ChatGPT hoe het ook weer zat met elliptische krommen en hun groepswet [3]. ChatGPT gaf een keurige uitleg, en kon op het eerste gezicht ook een voorbeeld met kleine gehele getallen geven. Alleen: daar klopte bij nadere beschouwing niets van.

Intussen weten we dat zulke ogenschijnlijk coherente verzinsels ('hallucinaties') schering en inslag zijn bij grote taalmodellen zoals ChatGPT. Er zijn vooralsnog geen bevredigende methoden beschikbaar voor goede controle. Sleutelen aan de wiskundige grondslagen is onontbeerlijk.



Wiskunde heeft vanaf het begin een belangrijke rol gespeeld in de ontwikkeling van kunstmatige intelligentie. Toen ik aan mijn wiskundestudie begon (aan de UvA, in 1985), was net het boek 'Gödel, Escher, Bach' [4] van Douglas Hofstadter in Nederlandse vertaling verschenen. Het was de tijd dat de belangrijkste ontwikkelingen in de AI gebaseerd waren op wiskundige logica. De nadruk lag

op logisch correct redeneren en de representatie van kennis in logische formules. In zijn boek behandelt Hofstadter de onvolledigheidsstellingen van Kurt Gödel en suggereert een verband tussen de zelfverwijzende zinnen in Gödel's bewijzen en de lusstructuren in onze hersenen. In een later boek 'I am a strange loop' legt hij nog meer nadruk op zijn visie dat ons 'zelf', ons 'ik', onze 'geest', ontstaat in complexe systemen met zelfverwijzende lusstructuren.

In de jaren daarna kantelde langzaam het belang van de wiskundige logica voor AI. Een destijds niet al te succesvolle AI-techniek die door velen niet erg serieus werd genomen (want er was geen solide basis in de logica) werd geleidelijk aan ontwikkeld tot het werkpaard van de AI van vandaag: neurale netwerken. De opkomst van het internet zorgde voor heel veel data, en neurale netwerken bleken zo juist heel veel te kunnen.

Inmiddels is er veel aandacht voor de gebreken van AI gebaseerd op neurale netwerken. Neurale netwerken geven geen garantie op correcte antwoorden. Ze geven geen relevante redenen voor hun conclusies. En ze kunnen zich niet aan normatieve en ethische regels houden. Interessant genoeg gaat het om gebreken die juist niet kleven aan AI gebaseerd op logica.

De zoektocht is nu dan ook naar technieken die de sterke kanten van op logica-gebaseerde systemen verbinden met die van lerende systemen zoals neurale netwerken. Vaak worden dan Bayesiaanse netwerken genoemd. Dat zijn wiskundige modellen die in de AI populair zijn omdat ze een compacte representatie van kansdistributies over veel variabelen mogelijk maken, en zo data compact kunnen modelleren. De onderliggende gerichte graaf van een Bayesiaans netwerk moet lusloos zijn en heeft wel wat weg van een causale structuur. In de praktijk blijkt het nog knap lastig om Bayesiaanse netwerken te gebruiken om de brug te slaan tussen logica-gebaseerde en lerende systemen (en causaliteit is al helemaal een verhaal apart).

We moeten dus nog geschiktere wiskunde verzinnen. Waar moeten we die zoeken? In Groningen, aan het Bernoulli instituut, wordt daarvoor aan argumentatiewiskunde gewerkt. Zo'n aanpak is relevant voor een recente ontwikkeling: hybride intelligente systemen. Dat zijn systemen waarin mens en machine samenwerken om elkaars sterke eigenschappen te versterken (in Nederland gesteund door het NWO Zwaartekracht project Hybrid Intelligence [5]). Die samenwerking kan geanalyseerd worden als een discussie tussen gesprekspartners die elkaar kritisch bevragen om problemen op te lossen. Oftewel: argumentatie.

De theorie van argumentatie gaat ver terug in de geschiedenis en is sterk verbonden met de ontwikkeling van logica en kanstheorie. Een tijdlang bewoog de argumentatietheorie zich juist weg van de wiskunde op zoek naar meer praktische relevantie. Verrassend genoeg keerde de argumentatiewiskunde terug in de vorm van grafentheorie. De knopen van een gerichte graaf worden geïnterpreteerd als abstracte argumenten en de pijlen geven aan welke argumenten andere argumenten weerleggen. Het blijkt dat lusstructuren een speciale plaats innemen: lussen van even lengte representeren verschillende coherente standpunten en lussen van oneven lengte een onopgelost conflict.

Terug naar de goede wiskunde voor de AI van vandaag. Eén idee waar in Groningen aan wordt gewerkt is om logica, kanstheorie en argumentatie wiskundig zuiver te combineren. Daarbij wordt logica gebruikt voor correct

redeneren met complexe kennis, kanstheorie voor het analyseren van patronen in data, en argumentatie voor het slaan van de brug tussen kennis en data. Een andere open puzzel is hoe discrete en continue AI methoden op elkaar inwerken. Terugkoppelingslusen in dynamische systemen lijken een relevante rol te spelen.

Op die manier krijgen we langzaam meer grip op de samenwerking tussen mens en machine. De mens is ontontbeerlijk in die vreemde lus want alleen mensen weten wat belangrijk is voor mensen. Machines kunnen dat niet zonder ons weten.



Torus met drie gaten als een vreemde lus in de stijl van Escher's waterval (MidJourney, 16 februari 2024)

References

- [1] <https://sites.google.com/view/wintersymposium2023/voordrachten-sprekers?authuser=0#h.yhdsaayicnzo>
- [2] <https://www.ai.rug.nl/~verheij/publications/uva1991.htm>
- [3] https://en.wikipedia.org/wiki/Elliptic_curve
- [4] https://en.wikipedia.org/wiki/Gödel,_Escher,_Bach
- [5] <https://www.hybrid-intelligence-centre.nl/research/>